13. ročník
Žilina 1. – 3. 7. 2025
Otvorený softvér vo vzdelávaní, výskume a v IT riešeniach
ossconf.fri.uniza.sk
OSSConf

OSSConf 2025: 1–3

LaTeX a unicode znaky

JOZEF ŘÍHA (SK) A PAVEL STŘÍŽ (CZ)

Abstrakt. Článek zmiňuje základní řešení, jak jít na sazbu speciálních znaků v LaTeXu a přímý zápis znaků v kódování UTF-8.

Klíčová slova. LaTeX, Ti*k*Z.

LaTeX AND THE UNICODE GLYPHS

Abstract. The paper introduces basic solution how to typeset special LaTeX symbols and use characters in UTF-8 encoding directly.

Keywords. LaTeX, Ti*k*Z.

1. Sazba štítků

Sázíme si štítky na součástky. Uložené je máme v `https://homebox.software/en/`. Zkusili jsme postupně následující.

- **gLabels:** Samolepící papír je rozdělený na 4 části a nechceme mít nálepku rozdělenou na 2 části.
- **Scribus:** Nenašli jsme cestu, jak škálovat písmo podle oblasti.
- **Mailmerge:** řešený zvláštně přes doplněk a i tak to není ono.
- Přišel čas oprášit **LaTeX.**

Sazba samotná není podstatná, zde jsou klíčové body.

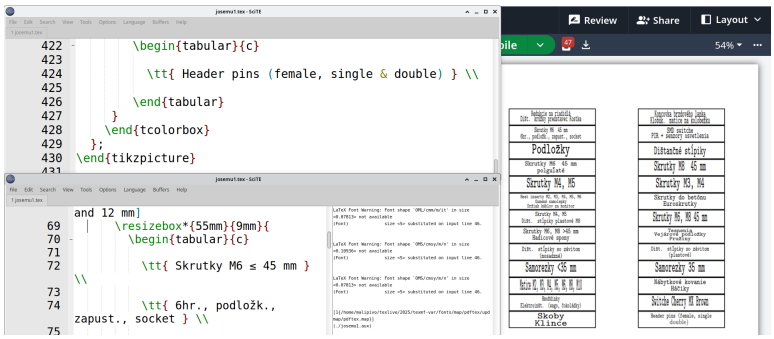
- Balíček tcolorbox s volitelným parametrem `fitting`.
- Prostředí `tcolorbox` s volitelným parametrem `fit to`.
- Balíček Ti*k*Z.
- Prostředí `tikzpicture` s parametry `remember picture` a `overlay`.
- Příkaz `\resizebox` z balíčku graphicx.

Náš cíl je neužít TeXové sekvence, ale mít je zapsané přímo. O balíčku tcolorbox má příspěvek Aleš Kozubík ve sborníku OSSConf z roku 2024.

2. Overleaf

Po běhu LaTeXu jsme nahlédli do log souboru:

- Znak & je aktivní znak, Overleaf používá `--interaction=nonstopmode`, vygeneruje PDF, ale chybu přeskočí.



- Znak ≤ patří mezi matematické symboly, zde je zapsán jako UTF-8 symbol v textovém režimu. Objeví se chyba: Missing character: There is no ≥ (U+2265) in font [...].

3. Řešení přes balíček newunicodechar

Na tomto místě si nebudeme ukazovat řešení přes XeLaTeX, kde lze skupinám znaků nadefinovat konkrétní písmo. Nebudeme ani poukazovat na experimentální balíček unicode-math. Poukážeme na méně známý balíček newunicodechar. Takto by vypadalo řešení po načtení balíčku v preambuli dokumentu.

```
\newunicodechar{≤}{\ifmmode\leq\else≤\fi}
\newunicodechar{≥}{\ifmmode\geq\else≥\fi}
```

4. Řešení přímé

Na první problém můžeme použít příkazy PlainTeXu. Inspiraci najdeme v knize *TEXbook naruby* od Petra Olšáka, `https://petr.olsak.net/ftp/olsak/tbn/tbn.pdf`. Přepneme si kategorii znaku & na 12 a bude se nám hezky sázet.

U UTF-8 znaků můžeme použít obrácený postup. Kategorii znaku si přepneme na aktivní. To je kategorie 13. A nadefinujeme si jej jako TeXový příkaz.

```
% ASCII kód & je 46_8, 38_10 a 26_16.

%\catcode'46=12
%\catcode38=12
\catcode"26=12

%\catcode"2264=13\def≤{$\leq$}
%\catcode"2265=13\def≥{$\geq$}

\catcode`≤=13\def≤{$\leq$}
\catcode`≥=13\def≥{$\geq$}
```

5. Zobecnění, vlastní styl

Pokud se nám zalíbila tato přímá cesta, pak by to chtělo jen rozšířit na větší škálu znaků. Sepíšeme několik tipů na zdroje.

- Za pokus by stálo nahlédnout na konverzní nástroje TeX na HTML.
- Ze zdrojových kódů dokumentů typu `https://www.math.uci.edu/~xiangwen/pdf/LaTeX-Math-Symbols.pdf`.
- V projektu LaTeX2Nemeth, `https://myria.math.aegean.gr/braille/symbols-in-braille.pdf`, bychom je také nejspíš našli.
- Ideál je nejspíš již zmíněný balíček unicode-math, obrázek vlevo, viz `https://mirrors.nic.cz/tex-archive/macros/unicodetex/latex/unicode-math/unimath-symbols.pdf`.
- Zaujal nás projekt LaTeX-unicoder.vim, obrázek vpravo, viz `https://github.com/joom/latex-unicoder.vim/blob/master/autoload/unicoder.vim`.

usv	L	X	S	P	D	F	N	H	E	C	R	Macro	Description
u+02294	⊔	⊔	⊔	⊔	⊔	⊔	⊔	⊔	⊔	⊔	⊔	\sqcup(p)	square union
u+02295	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	\oplus(p)	plus sign in circle
u+02296	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊖	\ominus(p)	minus sign in circle
u+02297	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	\otimes(p)	multiply sign in circle
u+02298	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	\oslash(p)	solidus in circle
u+02299	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	\odot(p)	middle dot in circle
u+0229A	⊚	⊚	⊚	⊚	⊚		⊚	⊚	⊚	⊚	⊚	\circledcirc(a)	small circle in circle
u+0229B	⊛	⊛	⊛	⊛	⊛		⊛	⊛	⊛	⊛	⊛	\circledast(a)	asterisk in circle
u+0229C	⊜	⊜	⊜	⊜	⊜		⊜	⊜	⊜	⊜	⊜	\circledequal	equal in circle
u+0229D	⊝	⊝	⊝	⊝	⊝		⊝	⊝	⊝	⊝	⊝	\circleddash(a)	hyphen in circle
u+0229E	⊞	⊞	⊞	⊞	⊞	⊞	⊞	⊞	⊞	⊞	⊞	\boxplus(a)	plus sign in box
u+0229F	⊟	⊟	⊟	⊟	⊟	⊟	⊟	⊟	⊟	⊟	⊟	\boxminus(a)	minus sign in box
u+022A0	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	⊠	\boxtimes(a)	multiply sign in box
u+022A1	⊡	⊡	⊡	⊡	⊡	⊡	⊡	⊡	⊡	⊡	⊡	\boxdot(a)	/dotsquare /boxdot b: small dot in box

latex-unicoder.vim / autoload / unicoder.vim

Code Blame 2695 lines (2671 loc) · 111 KB

```
152     \ '\Chi'                : 'Χ',
153     \ '\Psi'                : 'Ψ',
154     \ '\Omega'              : 'Ω',
155     \ '\alpha'              : 'α',
156     \ '\beta'               : 'β',
157     \ '\gamma'              : 'γ',
158     \ '\delta'              : 'δ',
159     \ '\upepsilon'          : 'ε',
160     \ '\zeta'               : 'ζ',
161     \ '\eta'                : 'η',
162     \ '\theta'              : 'θ',
163     \ '\iota'               : 'ι',
164     \ '\kappa'              : 'κ',
165     \ '\lambda'             : 'λ',
```

Pokud máme znaky či jejich UTF-8 hodnoty a příslušný TeXový příkaz, pak stačí jednoduchý skript, např. v Pythonu, a jsme s úlohou hotovi.

Kontaktní adresy

Ing. Jozef Říha, IBM Certified Specialist: pSeries Administration and Support, Bratislava, Slovenská republika,
E-mailová adresa: `jose1711@gmail.com`

Ing. Pavel Stříž, Ph.D., Nakladatelství Martin Stříž, U Škol 940, Bučovice, okres Vyškov, 685 01, Česká republika,
E-mailová adresa: `pavel@striz.cz`